

Paul J. Wolfangel

DATENVERARBEITUNG IM INSTITUT FÜR DEUTSCHE SPRACHE

1. Einleitung
2. Textverarbeitung
 - 2.1 Textbibliothek
 - 2.2 Möglichkeiten der Textaufbereitung
 - 2.3 Datenretrieval
 - 2.4 Statistik
 - 2.5 Parallelcodierung
3. Ausstattung mit Datenverarbeitungsgeräten
 - 3.1 Datenverarbeitungszentrum Mannheim (DVZ)
 - 3.2 IdS-Außenstelle Bonn
 - 3.3 IdS-Außenstelle Marburg
4. Programmierung
 - 4.1 Projekt LDV II
 - 4.1.1 Arbeitskomplexe
 - 4.2 Programme zur Textverarbeitung im DVZ
5. Service für Institutsmitglieder und Außenstehende
6. Kooperation
7. Projektträger
8. Bibliographische Anmerkungen

1. Einleitung

Über den Beginn der Aufnahme von Texten auf maschinell verarbeitbare Datenträger im Jahr 1965 und die Entwicklung der ersten Programme zur Auswertung dieser Texte bis zum Stand von 1968 haben Ingeborg Zint¹ und Manfred W. Hellmann² berichtet. Die Aufsätze enthalten eine Übersicht über das Mannheimer und Bonner Textcorpus, eine Diskussion der Fragestellungen an die Texte und die Beschreibung der Verfahren und Programme zur Textauswertung, wie sie im Deutschen Rechenzentrum in Darmstadt für Mannheim und im Bonner Universitätsrechenzentrum für die Bonner Außenstelle des Instituts mit Hilfe der Datenverarbeitungsanlage IBM 1410 und IBM 7090/94 entwickelt wurden.

Da die Arbeit an einem Fremdrechenzentrum die Möglichkeiten der Forschung und Entwicklung erheblich einschränkte, trat das Institut im Jahr 1967 in Verhandlungen über die Beschaffung einer eigenen Datenverarbeitungsanlage ein. Im Herbst 1969 wurden vom Bundesministerium für Bildung und Wissenschaft im Rahmen des Demonstrationsprojekts "Entwicklung eines Programmsystems für linguistische Aufgaben" die Mittel für eine Datenverarbeitungsanlage vom Typ Siemens 4004/35 bewilligt.

Kurz nach Installation dieser Anlage wurde die Projektteilung "Linguistische Datenverarbeitung" (LDV) gegründet. Die Arbeiten zum Programmsystem für linguistische Aufgaben konnten im Rahmen der Aufgabenstellung des Großprojekts ("Praxisorientierte Grundlagenforschung auf dem Gebiet der linguistischen Datenverarbeitung") weitergeführt werden.

Heute sind allein in der Abteilung LDV 28 Wissenschaftler und 28 weitere Mitarbeiter für Rechenzentrumsbetrieb und Verwaltung tätig. Hinzu kommen 8 über Werkvertrag beschäftigte Mitarbeiter für die System-, Grundlagen- und Anwendungsprogrammierung und ungefähr 30 wissenschaftliche und studentische Hilfskräfte.

Nicht nur die Abteilung LDV jedoch bedient sich der Datenverarbeitungsanlage als Forschungsinstrument und Hilfsmittel; auch Mitarbeiter aus anderen Abteilungen bzw. Projekten des Instituts, die das umfangreiche Textcorpus als Grundlage ihrer Forschung haben, benutzen die Aufbereitungs- und Retrieval-Programmsysteme, die im Laufe der letzten Jahre entwickelt wurden.

Daneben treffen in steigendem Maße Anfragen und Aufträge aus dem In- und Ausland ein. Zu ihrer Bearbeitung wurde inzwischen eine Clearing- und Servicestelle für deutsche Textverarbeitung eingerichtet.

Der folgende Bericht gibt einen Überblick über den Stand der Arbeiten und die gegenwärtig laufenden Aktivitäten.

2. Textverarbeitung

Eine ausführliche Dokumentation über die Textbibliothek, Konventionen zur Datenerfassung, Verarbeitungsprogramme und Ausgabemöglichkeiten erscheint in Kürze als Forschungsbericht³.

2.1 Textbibliothek

Zur Textbibliothek des Instituts gehören die sogenannten Bonner, Mannheimer und Freiburger Texte. Das Bonner Corpus, Texte aus den Tageszeitungen "Die Welt"

und "Neues Deutschland", hat derzeit einen Umfang von ca. 1,2 Millionen Wörtern. Das Mannheimer Corpus, eine Auswahl aus Dichtung, Trivialliteratur, wissenschaftlicher und populärwissenschaftlicher Literatur sowie Zeitungs- und Zeitschriftentexten, hat einen Umfang von ca. 1,8 Millionen Wörtern. Das Freiburger Corpus, eine auf Tonband erfaßte Sammlung weitgehend spontan gesprochener Texte, die für die Aufnahme auf maschinell verarbeitbare Datenträger transkribiert und unter der Leitung von Prof. Winkler, Marburg, mit Intonationsmerkmalen versehen wurden, umfaßt ungefähr 600.000 Wörter.

Erweiterungen der Textbasis sind in Bonn und Mannheim vorgesehen. Die Bonner Abteilung für Grammatik und Lexik will ihr Corpus von Zeitungstexten durch Artikel aus dem "Wuppertaler Generalanzeiger" und entsprechend aus Bezirkszeitungen der DDR ergänzen. In Mannheim arbeitet die Gruppe "Corpustheorie" an einer wissenschaftlich fundierten Konzeption für den Aufbau eines umfassenden Corpus, das dann u.a. auch Texte aus Illustrierten, Fachzeitschriften, Schulbüchern, sowie Geschäftskorrespondenz, Erlasse, Parteiprogramme und Gebrauchsanweisungen enthalten soll.

Alle vorhandenen Texte können in zeilennumerierter und satznumerierter Form ausgegeben und ausgedruckt werden. Groß-/Kleinschreibung ist möglich und wird in den meisten Fällen auch verwendet.

2.2 Möglichkeiten der Aufbereitung

Für Fragestellungen an den Wortschatz kann der gesamte Wortbestand der Texte alphabetisch oder nach Häufigkeit sortiert und als Wortform- oder Häufigkeitsre-

gister ausgegeben werden. Die Ausgabe kann auch mit der Option "restriktiv" oder "selektiv" erfolgen, d.h. es werden alle Wortformen außer den in einer vorgegebenen Liste enthaltenen bzw. nur die in einer vorgegebenen Liste enthaltenen Wortformen ausgedruckt. Soll die Sortierung über mehrere Texte hinweggehen, so wird für die Gesamtheit dieser Texte ein sogenanntes Gemischtes Register erstellt. Dabei wird festgehalten, aus welchem Text die Wortform stammt und in welcher Häufigkeit sie darin auftritt. Schließlich können Wortformen- und Gemischte Register auch rückläufig, d.h. von den Wortendungen her, sortiert werden; dies erlaubt einen alphabetischen Zugriff zu den Endungen.

Eine weitere Möglichkeit der Aufbereitung bietet ein Programm, das einen Schlüsselwortindex, ähnlich dem von IBM entwickelten KWIC ("Keyword-in-Context")-Index erstellt. Hier werden die alphabetisch sortierten Wortformen zentriert im Kontext je einer Zeile aufgelistet. Für die Bibliotheksdokumentation entwickelt, hat sich dieser Index auch als ein raum- und damit papiersparendes Hilfsmittel zur Untersuchung von Wortumgebungen erwiesen.

Auch hier sind selektive und restriktive Ausgabemöglichkeiten vorgesehen.

2.3 Retrieval

Die bisher geschilderten Aufbereitungsmöglichkeiten für die Texte ergeben sich im großen und ganzen aus der Umschichtung bzw. Sortierung der einzelnen Textelemente. Wenn bestimmte Begriffe (Buchstaben, Sonderzeichen, Wörter oder Kombinationen dieser Elemente) gesucht werden und der Ausdruck im Kontext eines vollständigen

Satzes oder mehrerer Sätze erwünscht ist, so kann dies mit einem dafür entwickelten Suchprogramm geschehen. Ausgedruckt werden die gefundenen Suchergebnisse entweder über die gesamte Breite des Druckpapiers, oder auf Wunsch im DIN A 6-, d.h. Karteikarten-Format. Es kann mit diesem Programm auch nach Zeichenfolgen gesucht werden, die im Text getrennt stehen.

2.4 Statistik

Die absolute und relative Häufigkeit des Auftretens bestimmter Merkmale kann durch ein Statistikprogramm abgefragt werden. Gezählt werden können Sätze, Zeilen, Wörter, Zeichen insgesamt, Sonderzeichen, Buchstaben, Vokale, Konsonanten und Zahlen. Bei jedem Statistiklauf werden außerdem die Extrem- und Durchschnittswerte für die in einem Satz vorkommende Zahl der Wörter und Zeichen und bei Bedarf die prozentualen Häufigkeiten für das Auftreten der einzelnen Buchstaben im Text ermittelt. Wird die Statistik nicht über den gesamten Text gewünscht, so kann über einen Zufallszahlengenerator anhand der Satznummern ein beliebiger Prozentsatz der im Text vorhandenen Sätze ermittelt und dann statistisch untersucht werden.

2.5 Parallelcodierung

Mit dem Parallelcodierungsverfahren (ein erster Entwurf ist im Forschungsbericht 2 des IdS, S. 30 ff. beschrieben⁴) können einem Text weitere Informationen zugeordnet werden, die als Ergebnisse manueller oder automatischer Analysevorgänge an diesem Text gewonnen wurden. Der Text kann auf beliebig vielen Analyseebenen nach beliebigen Kriterien segmentiert und die Segmente beliebigen Gesichtspunkten klassifiziert werden. Durch dieses Verfahren der Zuordnung kann der Originaltext

in seiner ursprünglichen Form verbleiben und muß nicht durch Dazwischenschieben von linguistischen, dokumentarischen oder anderen Codes "verunreinigt" werden. Eine Dokumentation des Verfahrens liegt vor und wird in einem der nächsten Forschungsberichte des IdS veröffentlicht.

3. Ausstattung mit Datenverarbeitungsgeräten

3.1 Datenverarbeitungszentrum Mannheim

Nach größeren Umbauarbeiten wurde im August 1974 ein neues Rechenzentrum in Betrieb genommen. Zwei Rechenanlagen sind zur Bearbeitung der vielfältigen Aufgaben von innerhalb und außerhalb des Instituts eingesetzt.

Die institutseigene Siemens-Datenverarbeitungsanlage vom Typ 4004/35 hat eine Speicherkapazität von 65 KB. Ihre Peripherie besteht aus vier 9-Spur-Magnetbandgeräten (30 KB/s) mit einer Aufzeichnungsdichte von 800 bpi, 2 Plattenspeichern mit einer Kapazität von je 7,25 Millionen Bytes und einer mittleren Zugriffszeit von ca. 90 Millisekunden. Der Schnelldrucker mit auswechselbarer Typenkassette hat 132 Schreibstellen und arbeitet wahlweise mit einem 64- oder einem 112-Zeichensatz (Groß- und Kleinbuchstaben sowie erweiterter Sonderzeichenvorrat); die maximale Druckgeschwindigkeit liegt bei ca. 55.000 Zeilen in der Stunde. Angeschlossen sind außerdem Geräte für die Lochkarten- sowie für die Lochstreifen-Ein-/Ausgabe. Der Lochstreifenleser verarbeitet 5-, 6-, 7- und 8-Kanal-Standardlochstreifen.

Zur Entwicklung und als Komponente eines experimentellen interaktiven Informationssystems wurde im Septem-

ber 1974 eine Großrechenanlage vom Typ Siemens 4004/151 auf Mietbasis installiert. Bei dieser Anlage steht dem Benutzer ein Adreßraum (virtueller Speicher) von über 8 Millionen Bytes (8000 KB) zur Verfügung, also ein Vielfaches des wirklich vorhandenen physikalischen Kernspeichers (384 KB).

Das System 4004/151 erlaubt mit Hilfe des Betriebssystems BS 2000 für virtuelle Speicher eine zeitlich ineinander verzahnte, gleichberechtigte Bearbeitung vieler verschiedener Programme. In Verbindung mit Datenübertragungseinrichtungen und Datenendgeräten ("Terminals") können Datenfernverarbeitung und Dialogverkehr durchgeführt werden, eine unabdingbare Voraussetzung für die Forschungsarbeit im Projekt LDV II.

Die Peripherie der Anlage besteht aus vier 9-Spur-Magnetbandgeräten (bis 240 KB/s) mit umschaltbarer Aufzeichnungsdichte 800/1600 bpi, 5 Plattenspeichern mit einer mittleren Zugriffszeit von 47,5 ms und einer Speicherkapazität von 55 Mio Bytes pro Plattenstapel.

Der Anlage 4004/151 kann bei Bedarf zu dem vorhandenen Schnelldrucker zusätzlich der Drucker des kleineren Rechners 4004/35 zugeordnet werden.

Ebenfalls kann der Lochkartenleser der 4004/35 zusätzlich zu dem an der 4004/151 betriebenen Lochkartenleser zugeschaltet werden.

4 Datensichtsstationen (mit einem Hardcopy-Gerät) in Mannheim und 2 Datenschreibstationen in Bonn dienen dem Mensch-Maschine-Dialog.

Für die Datenerfassung stehen in Mannheim 4 Karten-

locher (1 IBM 026, 1 IBM 029, 2 Siemens 2080-30), 2 Lochstreifen-Schreibautomaten und 2 Magnetband-Erfassungsgeräte Honeywell Keytape 900 zur Verfügung. 2 weitere Honeywell-Magnetband-Erfassungsgeräte sind in den Außenstellen des Instituts installiert; sie werden für die Datenerfassung eingesetzt und können zur Datenfernübertragung über das Telefonnetz an Mannheim angeschlossen werden. Das Bonner Gerät hat eine Lochkarten-Leseeinrichtung. Alle 4 Geräte können die erfaßten oder übertragenen Daten über eine gekoppelte elektrische Kugelkopf-Schreibmaschine ausdrucken.

Die Weiterverarbeitung des Drucker-Outputs der Datenverarbeitungsanlage (Separieren der Listen und Schneiden auf Format) erfolgt auf 2 Fimafold-Papierverarbeitungsmaschinen.

3.2 Außenstelle des Instituts in Bonn

Die LDV-Forschungsgruppe Bonn betreibt ihre Entwicklungen seit Installation der Datenschreibstationen größtenteils über Datenfernverarbeitung an der Mannheimer Großrechenanlage. Weitere Arbeiten, auch solche der Forschungsstelle für Grammatik und Lexik werden auf der IBM 370-168 am Universitätsrechenzentrum Bonn durchgeführt.

Zum Ausdrucken der dort beschriebenen Magnetbänder dient eine vom Institut gemietete programmierbare IBM Off-line-Druckstation 5960. Sie kann von 9-Kanal-Magnetbändern mit einer Zeichendichte von 800 oder 1600 bpi (auch in Groß-/Kleinschreibung) mit einer Geschwindigkeit von max. 36.000 Zeilen/Std. drucken.

2 IBM 026- und Siemens 2080-30-Kartenleser sowie

1 Supertyper-Schreibautomat zur Arbeit mit Lochstreifen stehen in Bonn zur Verfügung, außerdem das bereits erwähnte, auch zur Datenfernübertragung verwendbare Honeywell-Magnetband-Erfassungsgerät.

3.3 Außenstelle des Instituts in Marburg

Ein automatisches Zeichengerät der Firma AEG-Telefunken, Geograph 3000 Aristomat 8446, ist im Teilprojekt Sprachkartographie bei der Marburger Außenstelle eingesetzt. Das Zeichengerät arbeitet on-line an der Datenverarbeitungsanlage des Universitätszentrums Marburg, einer Telefunken TR4.

Ein Honeywell-Keytape für die Datenerfassung befindet sich ebenfalls in Marburg.

4. Programmierung

4.1 Projekt LDV II

Das Projekt LDV II in der Abteilung LDV setzt sich zum Ziel, ein experimentelles Informationssystem zu konzipieren und zu implementieren. Das Projekt läuft bis Ende 1975. Vorarbeiten dafür wurden im Projekt LDV I in den Jahren 1971 - 1973 geleistet.

Dieses in LISP zu programmierende Informationssystem auf linguistischer Basis (ISLIB) ist als ein automatisches, natürlichsprachlich fundiertes Frage-Antwort-System konzipiert. Es soll dementsprechend interaktiv sein und eine hohe Problemlösungskapazität aufweisen. Die Anforderung an ein solches System ist die Beantwortung natürlich-sprachlicher Fragen ebenfalls in natürlicher Sprache, selbst wo die Antwort nicht direkt in den vorhandenen Informationen zu finden,

sondern nur daraus erschließbar ist.

Die damit aufgeworfene Fragestellung kann als eine Schnittstelle von Teilproblematiken der Gesamtfor-
schungseinrichtung Artificial Intelligence ("Künstli-
che Intelligenz") betrachtet werden.

Die Relevanz dieses Gebiets liegt im Bestreben, zu einem gegebenen Ausgangspunkt und einem gewünschten Ziel (im vorliegenden Fall also einer Frage und einer Antwort) die richtigen Operationen, z.B. Ableitungsregeln, die beide miteinander verbinden, automatisch und ökonomisch zu finden, und dies möglichst in generalisierter Form zu programmieren, so daß nicht jedes einzelne Problem individuell vom Programmierer gelöst werden muß, sondern das System sich die Problemlösung aus den vorhandenen Bausteinen selbst aufbauen kann.

Das System soll interaktiv sein, d.h. es soll auf auftretende Schwierigkeiten mit gezielten Rückfragen und Beschreibung der Schwierigkeiten reagieren und neue Informationen des Benutzers akzeptieren und verwerten.

4.1.1 Arbeitskomplexe

Arbeitskomplex 1: Systementwicklung

Vorrangige Aufgabe ist die Festlegung der Systemsprache auf der Basis von LISP, in der das Experimentalsystem und seine Komponenten formuliert werden sowie die Konstruktion eines Prozessors, der diese Sprache verarbeiten kann. Sie ist mit hoher Priorität so zu konstruieren, daß die Verwendung der Systemsprache in allen Arbeitskomplexen möglich ist. Als Arbeitsschritte ergeben sich:

a) Festlegung der Systemsprache

- Explizierung der Systemsprache und Ausarbeitung einer geeigneten Terminologie
- Ausgrenzung von Subsprachen zur adäquaten Darstellung der unterschiedlichen innerhalb des Systems verwendeten oder zu erzeugenden Informationen
- hinreichende Definition der einzelnen Subsprachen in einer zeitlichen Reihenfolge, die auf die Bedürfnisse des gesamten Arbeitsplans Rücksicht nimmt
- Definition einer Verteilersprache, d.h. derjenigen Subsprache, in der Anweisungen zur Auswahl der jeweils erforderlichen Systemaktivität erfolgen

b) Programmierung des Experimentalsystems

- Implementierung einer sofort verfügbaren LISP-Version (Uppsala LISP-F2) als Übergangslösung für Schulungs- und Übungszwecke
- Konzipierung und Programmierung des Experimentalsystems als ablauffähiges Programmsystem (Steuerfeld)
- Programmierung der Ein- und Ausgabeverwaltung für die Interaktionsabwicklung (hierbei ist zur Zeit noch ungeklärt, welche Leistungen durch die ins Auge gefaßte LISP-Version erbracht werden und welche noch selbst programmiert werden müssen)
- Beschreibung und Programmierung der allgemeinen Formalismen der Datenbasis (Abspeichern und Aufsuchen von Daten bzw. Operationen aufgrund von Suchmustern)
- Algorithmen zur Einbettung der Elemente der Operationenbank in den Arbeitsablauf des Systems

Arbeitskomplex 2: Input-Sprache

Die Definition bzw. Konstruktion der Input-Sprache und damit auch zusammenhängend die Beschreibung der maschinell zu interpretierenden Eingabetexte erfolgt zum einen nach den in LDV I, vor allem durch die Gruppe "Maschinelle Syntaxanalyse" geleisteten Vorarbeiten, zum anderen ist sie abhängig von den in LDV II sich ergebenden notwendigen Erweiterungen. Diese beziehen sich zunächst auf Anforderungen, wie sie die Fallstudien an die kommunikative Kapazität und Ausdrucksfähigkeit der Input-Sprache stellen.

Arbeitskomplex 3: Fallstudien

Im Gegensatz zu Arbeitskomplex 4 "Problemfeldbehandlung" behandeln die ausgewählten Fallstudien "Börsenberichte" und "Grundwasserverschmutzung" konkrete Problemstellungen und bieten konkrete Problemlösungsverfahren an. Der zentrale Aspekt liegt bei der Untersuchung der Interaktionsstruktur und Interaktionskapazität beim Aufbau des ISLIB. Wegen dieser Ausrichtung auf realisierbare Interaktionsmechanismen muß detailliertes Fachwissen in die konkrete Problemstellung mit der dazugehörigen Empirie nicht eingebracht werden.

Arbeitskomplex 4: Problemfeldbehandlung

Während im Arbeitskomplex "Fallstudien" die Problemlösung im wesentlichen vom empirischen Material der Fallstudien aus angegangen wird, soll im Arbeitskomplex 4 die theoretische Aufhellung und Generalisierung der Problemtheorie vorangetrieben werden. Diese theoretischen Ergebnisse sind mit den Resultaten der empirischen Untersuchungen in den Fallstudien systematisch zu vereinigen, um die konstruktive Voraussetzung

für den Aufbau der problemorientierten Daten- und Operationenbank zu schaffen.

4.2 Programme zur Textverarbeitung im Datenverarbeitungszentrum (DVZ)

Die Programme zur Textverarbeitung auf der 4004/35 für das Mannheimer und Freiburger Corpus bewältigen folgende Aufgaben: Datenaufnahme und Wechsel von Datenträgern, Umcodieren, Korrektur, Zerlegen, Sortieren und Mischen, Aufbau von Indices und Lexika, statistische Ermittlungen und Ausdrucken in einer großen Auswahl von Formaten.

Die meisten Programme sind, u.a. aus Kompatibilitätsgründen, in FORTRAN IV geschrieben. Gelegentlich wird wegen Mangels an Speicherplatz oder zur Beschleunigung des Ablaufs auf Assembler ausgewichen. Einige wenige ALGOL-Programme werden zur Zeit in FORTRAN umgearbeitet.

5. Service für Institutsmitglieder und Außenstehende

Auf der kleineren Mannheimer Datenverarbeitungsanlage wurden in einem Zeitraum von 9 Monaten Service-Aufträge mit einem Volumen von 15,9 Millionen Wörtern Output bearbeitet. Die dafür aufgewendete Rechenzeit belief sich auf 319 Stunden, das sind über 35 Stunden im Monatsdurchschnitt.

Das Verhältnis von institutsinternen zu externen Aufträgen war in diesem Zeitraum fast genau 1:1 (160 Rechnerstunden und 7,9 Millionen Wörter Output für Mitarbeiter des Instituts, 159 Stunden und 8 Millionen Wörter Output für andere Forschungsinstitutionen des In- und Auslandes).

Aufträge von außerhalb erhielt das Datenverarbeitungs-

zentrum bisher von den Universitäten Berlin (FU und Akademie der Künste), Hamburg, München, Saarbrücken und Trier sowie aus dem Ausland von den Universitäten Kopenhagen, Lille, Manitoba, Princeton, Toulouse und Zagreb.

Im einzelnen handelte es sich dabei um Ausdrücke von Wortformen- und Häufigkeitsregistern, Ausdrücke satznumerierter und zeilennumerierter Texte, Suchläufe nach vorgegebenen Merkmalen (Wörter, Zeichenkombinationen, Codes) mit Ausdruck der gefundenen Sätze auf Karteikartenformat, Ausdrücke parallelcodierter Daten der geschriebenen und gesprochenen Sprache, Auflistung und Formatierung von Text- oder Code-Lochkarten, Sortierung von Daten und Textstatistiken. Gelegentlich wurden auch Lochkarten- und Lochstreifendaten der Auftraggeber aufgenommen und umcodiert, Magnetbanddaten kopiert und aufbereitet sowie Fremddaten an die Mannheimer Konventionen angepaßt.

Nicht in Zahlen fassen läßt sich die Informationsarbeit, die Beratung von Benutzern bei der Datenerfassung und -organisation, die Aufklärung über die Möglichkeiten der Auswertung des Textcorpus oder die auf Anregung von Benutzern vorgenommene Erweiterung unserer Programmbibliothek.

6. Kooperation

Bereits seit längerer Zeit wird mit universitären Forschungsgruppen zusammengearbeitet, die eine ähnliche Zielsetzung wie die der Projektabteilung LDV haben. Es bestehen auch Kontakte zu den Forschungsgruppen CONDOR bei Siemens und TELDOK bei Telefunken-Computer sowie zum Wissenschaftlichen Zentrum der IBM in Heidelberg.

Durch Teilnahme an wissenschaftlichen Tagungen, besonders auch an den Sitzungen von "LDV-Fittings", soll eine überregionale Koordinierung der Bemühungen auf dem Gebiet der Linguistischen Datenverarbeitung angestrebt werden. Über die Service- und Clearingsstelle der Abteilung ist es außerdem gelungen, einen Austausch von Daten und Informationen mit ausländischen Forschungsgruppen in die Wege zu leiten und die Arbeiten des Instituts im Datenverarbeitungsbereich der internationalen Fachwelt vorzustellen.

7. Projektträger

Die Gesellschaft für Mathematik und Datenverarbeitung (GMD) hat ab 1. September 1974 die Projektträgerschaft für alle Förderbereiche übernommen, die sie durch ihre Abteilung für Projektmanagement vorher nur in Form der Projektbegleitung betreut hat. Die GMD wird somit auch den Fortschritt des Projekts LDV überwachen und dem Bundesminister für Forschung und Technologie berichten.

8. Bibliographische Anmerkungen

- 1 Zint, Ingeborg: Maschinelle Sprachbearbeitung des Instituts für deutsche Sprache in Mannheim. - In: Forschungsbericht 2 des Instituts für deutsche Sprache, Mannheim 1968.
- 2 Hellmann, Manfred W.: Zur Dokumentation und maschinellen Bearbeitung von Zeitungstexten in der Außenstelle Bonn. - In: Forschungsbericht 2 des Instituts für deutsche Sprache, Mannheim 1968.
- 3 Bayer, Klaus/Epp, Berthold/Kurbel, Karl: Maschinelle Textverarbeitung im Rechenzentrum des Instituts für deutsche Sprache in Mannheim. - In: Forschungsbericht des Instituts für deutsche Sprache, Mannheim 1974.
- 4 Ströbl, Alex: Das Verfahren "Parallelcodierung". - Bei: Zint, Ingeborg: Maschinelle Sprachbearbeitung des Instituts für deutsche Sprache in Mannheim. - In: Forschungsbericht 2 des Instituts für deutsche Sprache, Mannheim 1968.